

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:21:42 ; Search time 24.98 Seconds

(without alignments)  
173.133 Million cell updates/sec

Title: US-09-786-214-5

Perfect score: 130

Sequence: 1 MAGLPAVGLSPGEQEHKRGCVGL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archae: \*  
2: SP-bacteria: \*  
3: SP-fungi: \*  
4: SP-human: \*  
5: SP-invertebrate: \*  
6: SP-mammal: \*  
7: SP-mhc: \*  
8: SP-organelle: \*  
9: SP-phage: \*  
10: SP-plant: \*  
11: SP-todent: \*  
12: SP-virus: \*  
13: SP-vertebrate: \*  
14: SP-unclassified: \*  
15: SP-rvirus: \*  
16: SP-bacteriap: \*  
17: SP-archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	39.2	426	17	Q9YBR0 aeropyrum p
2	50	38.5	277	16	Q9KQK6 vibrio chol
3	50	38.5	381	16	Q9KRL5 delinococcus
4	49.5	38.1	192	15	Q9GSR2 human immun
5	49.5	38.1	192	15	Q9GSR2 human immun
6	49	37.7	90	15	Q9GQK2 human immun
7	49	37.7	90	15	Q9GQK2 human immun
8	49	37.7	90	15	Q9GQK1 human immun
9	49	37.7	90	15	Q9GQK0 human immun
10	49	37.7	363	13	Q9DDH0 neoceratodu
11	49	37.7	4823	13	Q9G321 fugu rubrip
12	48.5	37.3	192	15	Q9LW51 human immun
13	48.5	37.3	192	15	Q9LW51 human immun
14	48.5	37.3	192	15	Q9LW51 human immun
15	48.5	37.3	195	3	Q9USC7 schizosacch
16	48	36.9	192	15	Q9GQ29 human immun

17	48	36.9	278	16	Q9CMF6 pasteurilla
18	48	36.9	792	16	Q9PEZ4 xylella fas
19	47.5	36.5	192	15	Q9WP98 human immun
20	47.5	36.5	192	15	Q9G4N4 human immun
21	47.5	36.5	192	15	Q9G4N4 human immun
22	47.5	36.5	192	15	Q9G4N4 human immun
23	47.5	36.5	192	15	Q9G4N4 human immun
24	47	36.2	192	15	Q9G4B2 human immun
25	47	36.2	293	16	Q9G4B2 human immun
26	47	36.2	293	16	Q9G4B2 human immun
27	47	36.2	390	16	Q9G4B2 human immun
28	47	36.2	661	10	Q9G4B2 human immun
29	47	36.2	813	11	Q9G4B2 human immun
30	47	36.2	2109	13	Q9G4B2 human immun
31	46.5	35.8	178	16	Q9HFK1 pseudomonas
32	46.5	35.8	192	15	Q9WP96 human immun
33	46.5	35.8	192	15	Q9G4B2 human immun
34	46.5	35.8	192	15	Q9G4B2 human immun
35	46.5	35.8	192	15	Q9G4B2 human immun
36	46.5	35.8	192	15	Q9G4B2 human immun
37	46.5	35.8	192	15	Q9G4B2 human immun
38	46.5	35.8	563	2	Q9X8S9 streptomyces
39	46.5	35.8	601	16	Q9X8B1 chlamydia m
40	46.5	35.8	656	16	Q9X8B1 chlamydia m
41	46.5	35.8	745	2	Q9L0S2 streptomyces
42	46	35.4	184	16	Q9A162 streptomyces
43	46	35.4	234	2	Q9F0S9 thermus the
44	46	35.4	271	16	Q9PG20 xylella fas
45	46	35.4	274	16	Q9PPC9 campylobact

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	426 AA.
Q9YBR0	Q9YBR0			
AC	Q9YBR0			
DT	01-NOV-1999 (TREMURel. 12, Created)			
DT	01-NOV-1999 (TREMURel. 12, Last sequence update)			
DT	01-JUN-2001 (TREMURel. 17, Last annotation update)			
DE	HYPOTHETICAL 48.5 KDA PROTEIN APE1539.			
GN	APE1539.			
OS	Aeropyrum pernix.			
OC	Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;			
CC	Aeropyrum.			
OX	NCBI_TaxID=56636;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-K1;			
EX	MEDLINE=99310339; PubMed=10382966;			
RA	Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,			
RA	Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankel A., Kosugi H.,			
RA	Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,			
RA	Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,			
RA	Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,			
RA	Nakamura Y., Nomura N., Sako Y., Kikuchi H.,			
RT	"Complete genome sequence of an aerobic hyper-thermophilic			
RT	crenarchaeon, Aeropyrum pernix K1."			
RL	DNA Res. 6:83-101(1999).			
DR	EMBL: AP000061; BAA80538.1; -			
KW	HYPOTHETICAL protein; Complete proteome.			
SO	SEQUENCE 426 AA; 48461 MW; 6D3D3469392D4DF0 CRC64;			

QY	3	GLPAVGLSPGEQEHKRGCV 24	39.2%; Score 51; DB 17; Length 426;
Db	278	GLADLIGEPGRERREGRV 299	Best Local Similarity 45.5%; Pred. No. 31;
			Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```

RESULT 2
O9KOK6 PRELIMINARY; PRT: 277 AA.
AC O9KOK6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE FORMYLETERRAHYDROFOLATE DEFORMYLASE.
GN VCL192.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N1661 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Enoliava M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae."
RL Nature 406:477-483(2000).
DR EMBL: AE004274; AAF95140.1; -
DR HSSP: P08179; ZGAR.
DR TIGR: VC1992; -
DR InterPro: IPR002912; ACT.
DR Pfam: PF01842; ACT; 1.
KM Complete proteome.
SQ SEQUENCE 277 AA; 31373 MW; A703491654753DC6 CRC64;

Query Match 38.5%; Score 50; DB 16; Length 277;
Best Local Similarity 52.9%; Pred. No. 26;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 4 LPAVGLSPGGEYHRG 20
DB 190 LPAFIGAKPYQOAYERG 206

RESULT 3
O9RRL5 PRELIMINARY; PRT: 381 AA.
AC O9RRL5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CYTOCHROME P450.
GN DR2473.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uitterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1."
RL Science 286:1571-1577(1999).
CC -I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AE002076; AAF12016.1; -.

```

```

DR TIGR: DR2473; -.
DR InterPro: IPR001128; Cyt_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN.1.
KM Complete proteome; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 381 AA; 41940 MW; F191EA69F1797B53 CRC64;

Query Match 38.5%; Score 50; DB 16; Length 381;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GLPAVGLSP 12
DB 51 GLPAVGLSP 60

RESULT 4
O9QSR2 PRELIMINARY; PRT: 192 AA.
AC O9QSR2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VIF PROTEIN.
GN VIF.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VI850;
RX MEDLINE=20192166; PubMed=10725202;
RA Laukkanen T., Carr J.K., Janssens W., Liltsola K., Gotte D.,
RA McCutchan F.E., Op de Coul E., Cornelissen M., Heyndrickx L.,
RA van der Groen G., Salminen M.O.;
RT "Virtually full-length subtype F and F/D recombinant HIV-1 from Africa
and South America."
RL Virology 269:95-104(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VI850;
RA Carr J.K., Kim B., Sanders-Buell E., Salminen M.O., Alaeus A.,
RA Albert J.A., Birx D.L., McCutchan F.E.;
RT "HIV-1 isolate VI850 from Zaïre, complete genome."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF077336; AAD46089.1; -
DR InterPro: IPR000475; Viral_infect.
DR Pfam: PF00559; VIF; 1.
DR PRINTS: PR00349; VIRIONINFECT.
DR ProDom: PD000063; Viral_infect; 1.
SQ SEQUENCE 192 AA; 22629 MW; F165805BPCDA4427 CRC64;

Query Match 38.1%; Score 49.5; DB 15; Length 192;
Best Local Similarity 58.8%; Pred. No. 21;
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

OY 9 GLSPGGEYHRG-GYGV 24
DB 71 GLHPRERHHLQGVSI 87

RESULT 5
O994O2 PRELIMINARY; PRT: 192 AA.
AC O994O2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE VIF PROTEIN.
GN VIF.

```

OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=96ZM751;  
RX MEDLINE=21094715; PubMed=11177395;  
RA Rodenburg C.M., Li Y., Traak S.A., Chen Y., Decker J., Robertson D.L.,  
RT Kallish M.L., Shaw G.M., Allen S., Hahn B.H., Gao F.;  
RT "Near full-length clones and reference sequences for subtype C  
RT isolates for HIV type 1 from three different continents.";  
RT AIDS Res. Hum. Retroviruses 17:161-168(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=96ZM751;  
RA Rodenburg C.M., Li Y., Traak S.A., Chen Y., Decker J., Robertson D.L.,  
RT Allen S., Shaw G.M., Hahn B.H., Gao F.;  
RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF286225; AAK30974.1;  
DR InterPro: IPR000475; Viral\_infect.  
DR Pfam: PF00559; Vif, 1.  
DR PRINTS: PR00349; VIRIONINFECT.  
DR ProDom: PD000063; Viral\_infect. 1.  
SQ SEQUENCE 192 AA; 22727 MW; 55E01D4BBD93DC6 CRC64;

Query Match 38.1%; Score 49.5; DB 15; Length 192;  
Best Local Similarity 58.8%; Pred. No. 21;  
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
OY 9 GLSPGGEYHRC-GVGV 24  
DB 71 GLHGEREMHGHGYSI 87

RESULT 6  
O9QOA2 PRELIMINARY; PRT; 90 AA.  
AC O9QOA2;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ENVELOPE PROTEIN (FRAGMENT).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 074;  
RX MEDLINE=99376522; PubMed=10449282;  
RA Abebe A., Demissie D., Goudsmit J., Brouwer M., Kuiken C.L.,  
RT Pollakis G., Schultemake H., Fontanet A.L., Rinke de Wit T.F.;  
RT "HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and  
RT coreceptor usage among Ethiopian patients with AIDS.";  
RL AIDS 13:1305-1311(1999).  
DR EMBL; AF158901; AAD46328.1;  
DR InterPro: IPR000777; GP120.  
DR Pfam: PF00516; GP120, 1.  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1  
FT NON\_TER 90  
SQ SEQUENCE 90 AA; 10453 MW; 8136F016865AADFA CRC64;

Query Match 37.7%; Score 49; DB 15; Length 90;  
Best Local Similarity 43.8%; Pred. No. 10;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
OY 8 VGLSPGGEYHRCGVG 23  
DB 38 IGIGPGQAFARGIG 53

RESULT 7  
O9QOA1 PRELIMINARY; PRT; 90 AA.  
AC O9QOA1;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ENVELOPE PROTEIN (FRAGMENT).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 074;  
RX MEDLINE=99376522; PubMed=10449282;  
RA Abebe A., Demissie D., Goudsmit J., Brouwer M., Kuiken C.L.,  
RT Pollakis G., Schultemake H., Fontanet A.L., Rinke de Wit T.F.;  
RT "HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and  
RT coreceptor usage among Ethiopian patients with AIDS.";  
RL AIDS 13:1305-1311(1999).  
DR EMBL; AF158901; AAD46328.1;  
DR InterPro: IPR000777; GP120.  
DR Pfam: PF00516; GP120, 1.  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1  
FT NON\_TER 90  
SQ SEQUENCE 90 AA; 10393 MW; 6073E4538641B74B CRC64;

Query Match 37.7%; Score 49; DB 15; Length 90;  
Best Local Similarity 43.8%; Pred. No. 10;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
OY 8 VGLSPGGEYHRCGVG 23  
DB 38 IGIGPGQAFARGIG 53

RESULT 8  
O9QOA0 PRELIMINARY; PRT; 90 AA.  
AC O9QOA0;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE ENVELOPE PROTEIN (FRAGMENT).  
GN ENV.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 074;  
RX MEDLINE=99376522; PubMed=10449282;  
RA Abebe A., Demissie D., Goudsmit J., Brouwer M., Kuiken C.L.,  
RT Pollakis G., Schultemake H., Fontanet A.L., Rinke de Wit T.F.;  
RT "HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and  
RT coreceptor usage among Ethiopian patients with AIDS.";  
RL AIDS 13:1305-1311(1999).  
DR EMBL; AF158902; AAD46329.1;  
DR InterPro: IPR000777; GP120.  
DR Pfam: PF00516; GP120, 1.  
KW AIDS; Coat protein; Glycoprotein.  
FT NON\_TER 1  
FT NON\_TER 90  
SQ SEQUENCE 90 AA; 10393 MW; 6073E354F631B74B CRC64;

Query Match 37.7%; Score 49; DB 15; Length 90;  
Best Local Similarity 43.8%; Pred. No. 10;  
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 8 VGLSPGEQEHYHRCVG 23  
 : : : : :  
 Db 38 IGIGPGQAFPARGGIG 53

RESULT 9  
 ID 090096 PRELIMINARY; PRT; 90 AA.  
 AC 090096;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ENVELOPE PROTEIN (FRAGMENT).  
 GN ENV.  
 OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_Taxid=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PATIENT 074;  
 RX MEDLINE=9376522; PubMed=10449282;  
 RA Abebe A., Demissie D., Goudsmit J., Brouwer M., Kuiken C.L.,  
 RA Pollakis G., Schuitemaker H., Fontanet A.L., Rinke de Wit T.F.;  
 RT "HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and  
 RT coreceptor usage among Ethiopian patients with AIDS.";  
 RL AIDS 13:1305-1311(1999).  
 DR EMBL; AF158906; AAD46333.1; -  
 DR InterPro; IPR000777; GFI20.  
 DR Pfam; PF00516; GFI20.1.  
 DR AIDS; Coat protein; Glycoprotein.  
 FT NON\_TER 1  
 FT 90  
 FT SEQUENCE 90 AA; 10377 MW; 9077E754EB2CA4D6 CRC64;

Query Match 37.7%; Score 49; DB 15; Length 90;  
 Best Local Similarity 43.8%; Pred. No. 10;  
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 8 VGLSPGEQEHYHRCVG 23  
 : : : : :  
 Db 38 IGIGPGQAFPARGGIG 53

RESULT 10  
 ID 09DDH0 PRELIMINARY; PRT; 363 AA.  
 AC 09DDH0;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ENOLASE (EC 4.2.1.11) (2-PHOSPHOGLYCERATE DEHYDRATASE) (2-PHOSPHO-D-  
 DE GLYCERATE HYDRO-LYASE) (FRAGMENT).  
 OS Neocercatodus forsteri (Australian lungfish).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Dipnoi; Ceratodontiformes; Ceratodontidae; Neocercatodus.  
 OX NCBI\_Taxid=7892;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21100378; PubMed=11163970;  
 RA Tracy M.R., Hedges S.B.;  
 RT "Evolutionary history of the enolase gene family.";  
 RL Gene 259:129-138(2000).  
 CC -1- CATALYTIC ACTIVITY: 2-PHOSPHO-D-GLYCERATE = PHOSPHOENOLPYRUVATE +  
 CC H(2)O.  
 CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING  
 CC THE DIMER (BY SIMILARITY).  
 CC -1- PATHWAY: GLYCOLYSIS.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.  
 DR EMBL; AY005158; AAG16307.1; -.

DR HSP; P56252; 1PDZ.  
 DR InterPro; IPR000941; Enolase.  
 DR Pfam; PF00113; enolase; 1.  
 DR PRINTS; PR00148; ENOLASE.  
 DR PRODOM; PD000902; ENOLASE; 1.  
 DR PROSITE; PS00164; ENOLASE; 1.  
 KW Glycolysis; Lyase; Magnesium.  
 FT NON\_TER 1  
 FT 363  
 FT SEQUENCE 363 AA; 39309 MW; 8049DB5E46EDA092 CRC64;

QY 2 AGVGLSPGEQEHYHRCVG 21  
 : : : : :  
 Db 209 AGVGLSPGEQEHYHRCVG 230

RESULT 11  
 ID 093321 PRELIMINARY; PRT; 4823 AA.  
 AC 093321;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ALL-1 RELATED PROTEIN.  
 GN ALR.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Takifugu.  
 OX NCBI\_Taxid=31033;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9177347; PubMed=10077531;  
 RA Gellner K., Brenner S.;  
 RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu  
 RT rubripes.";  
 RL Genome Res. 9:251-258(1999).  
 CC -1- SIMILARITY: CONSTRAINTS 1 SET DOMAIN.  
 CC EMBL; AF056116; AAC34383.1; -  
 DR InterPro; IPR003889; FYRICH\_C.  
 DR InterPro; IPR003888; FYRICH\_N.  
 DR InterPro; IPR000910; HMG\_12\_box.  
 DR InterPro; IPR001965; PHD.  
 DR InterPro; IPR003616; PostSET.  
 DR InterPro; IPR001214; SET.  
 DR InterPro; IPR001841; Znf\_ring.  
 DR Pfam; PF00628; PHD; 6.  
 DR Pfam; PF00856; SET; 1.  
 DR SMART; SM00542; FYRC; 1.  
 DR SMART; SM00541; FYRN; 1.  
 DR SMART; SM00398; HMG; 1.  
 DR SMART; SM00249; PHD; 8.  
 DR SMART; SM00508; PostSET; 1.  
 DR SMART; SM00184; RING; 4.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PS00280; SET; 2.  
 DR SEQUENCE 4823 AA; 526260 MW; BD0C5FAEAD0F9C7 CRC64;

Query Match 37.7%; Score 49; DB 13; Length 4823;  
 Best Local Similarity 66.7%; Pred. No. 9.3e+02;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 PAVVGLSPGEQEHYHRCVG 19  
 : : : : :  
 Db 1900 PALGLSPSELKRR 1914

ID	070887		PRT:	192 AA.
AC	070887			
DT	01-AUG-1998	(TREMBLrel. 07, Created)		
DR	01-AUG-1998	(TREMBLrel. 07, Last sequence update)		
DJ	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	VIF..			
GN	VIF..			
OS	Human immunodeficiency virus type 1.			
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.			
OX	NCBI_TaxID=11676;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=93BR020;			
RX	MEDLINE=98285725; PubMed=9621027;			
RA	Gao F., Robertson D.L., Carruthers C.D., Morrison S.G., Jian B., Chen Y., Barre-Sinoussi F., Girard M., Srinivasan A., Abimiku A.G., Shaw G.M., Sharp P.M., Hahn B.H.;			
RT	"A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1.";			
RT	J. Virol. 72:5680-5698(1998).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=93BR020;			
RA	Gao F.;			
RL	Submitted (MAY-1997) to the EMBL/GenBank/DBIJ databases.			
DR	EMBL: AF005494; AAD03170.1; -			
DR	InterPro: IPR000475; Vifrel_Infec.			
DR	Pfam: PF00559; Vif: 1.			
DR	PRINTS: PR00349; VIRIONINFECT			
DR	Prodont: PD000063; Vifrel_Infec: 1.			
SO	SEQUENCE 192 AA; 22641 MW; 53AEI79683DC137E CRC64;			

Query Match	37.3%; Score 48.5; DB 15; Length 192;
Best Local Similarity	58.8%; Pred. No. 29;
Matches 10; Conservative	3; Mismatches 3; Indels 1; Gaps 1

QY	9 GLSPGEQETHRG-GYGV 24
	:
Db	71 GLPGERMHGQGVSI 87

RESULT 13	
Q9IW51	
AC	Q9IW51; PRELIMINARY; PRT: 192 AA.
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DJ	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DJ	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	VIF PROTEIN.
GN	VIF..
OS	Human immunodeficiency virus type 1.
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX	NCBI_TaxID=11676;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=98SE-MP1211;
RA	Montavon C.;
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBIJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=98SE-MP1211;
RX	MEDLINE=20236909; PubMed=10777151;
RA	Toure-Kane C., Montavon C., Paye M., Gueye P., Sow P., Noye I., Gaye-Diallo A., Delaporte E., Peeters M., Mboup S.;
RT	"Identification of all HIV type 1 group M subtypes in Senegal, a country with low and stable seroprevalence.";
RL	AIDS Res. Hum. Retroviruses 16:603-609(2000).
RN	[3]

```

RP      SEQUENCE FROM N.A.
RC      STRAIN-98SE-MP1211.
RA      Montavon C., Touze-Fane C., Liegeois F., Mpoudi E., Bourgeois A.,
RA      Vergne L., Perret J.L., Boumah A., Saman E., Moup S., Delaporte E.,
RA      Peeters M.;
RT      "Most env and gag subtype A HIV-1 viruses circulating in West and
RT      West Central Africa are similar to the prototype AG recombinant virus
RT      IBNG.";
RL      J. Acquir. Immune Defic. Syndr. 15:363-374(2000).
DR      EMBL, AJ251056, CAB86367.1;-.
DR      InterPro, IPR000475; Viral_infect.
DR      Pfam, PF00559; Vif; 1.
DR      PRINTS; PR00349; VIRIONINFECT.
DR      PRODOM; PD000063; Viral_infect; 1.
SQ      SEQUENCE   192 AA;  22756 MW;  4FA1575561EC214 CRC64;

Query Match          37.3%; Score 48.5; DB 15; Length 192;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 9; Conservative 5; Mismatches 3; Indels 1; Gaps 1

Oy      9 GLSPGEQEHNG-GVGV 25
Db      71 GLQPERDWHLGHVSI 88

RESULT  14
O90CX5  PRELIMINARY; PRT; 192 AA.
ID      O90CX5.
AC      O90CX5.
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      VIF PROTEIN.
GN      VIF.
OS      Human immunodeficiency virus type 1.
OC      Viruses; Retroid viruses; Retroviridae; Lentivirns.
OX      NCBI_TaxID=11676;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-977207;
RX      MEDLINE=21395692; PubMed=11504977;
RA      Hoelscher M., Kim B., Maboko L., Mhalu F., von Sonnenburg F.,
RA      Birx D.L., McCutchan F.E.,
RA      the UNAIDS Network for HIV Isolation Characterization.;
RT      "High proportion of unrelated HIV-1 intersubtype recombinants in the
RT      Mbeja region of southwest Tanzania."
RT      AIDS 15:1461-1470(2001).
DR      EMBL, AF361877; AAK94266.1;-.
SQ      SEQUENCE   192 AA;  22715 MW;  3BB05BC5CB303B9D CRC64;

Query Match          37.3%; Score 48.5; DB 15; Length 192;
Best Local Similarity 58.8%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1

Oy      9 GLSPGEQEHNG-GVGV 24
Db      71 GLQPERDWHLGHVSI 87

RESULT  15
O9USC7  PRELIMINARY; PRT; 195 AA.
ID      O9USC7.
AC      O9USC7.
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      HYPOTHEICAL 22.1 KDA PROTEIN (FRAGMENT).
GN      SPAC23D3.12.
OS      Schizosaccharomyces pombe (Fission yeast).
OC      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC      Schizosaccharomycetales; Schizosaccharomycetaceae;
CC      CC

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:22:12 ; Search time 10.34 Seconds

(without alignments)  
93.616 Million cell updates/sec

Title: US-09-786-214-5

Sequence: 1 MAGLPAVGLSGPEGEYHRCGVCL 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	49	37.7	278	1	PURU_HAEN
2	48.5	37.3	559	1	Q03432 haemophilus
3	48	36.9	563	1	Q09852 schizosacch
4	48	36.9	780	1	P58543 versinia pe
5	47	36.2	280	1	Q03026 deinnococcus
6	47	36.2	280	1	PURU_ECOTI
7	47	36.2	2109	1	PURU_SHIFT
8	46	35.4	508	1	PECA_CHICK
9	45	34.6	479	1	COBI_MYCTU
10	45	34.6	961	1	PTSB_VIBAL
11	44	33.8	151	1	ACON_MYCAV
12	44	33.8	433	1	SODC_XIRGL
13	44	33.8	565	1	ENOB_RABIT
14	44	33.8	565	1	ARAB_ECO57
15	44	33.8	565	1	ARAB_ECOLI
16	44	33.8	617	1	ARAB_SALTI
17	44	33.8	658	1	ILVD_STRCO
18	44	33.8	677	1	VC18_BPT4
19	44	33.8	777	1	SGI_HUMAN
20	44	33.8	1065	1	L100_ADE41
21	43.5	33.5	968	1	RPOB_MARPO
22	43	33.1	221	1	CTDI_HUMAN
23	43	33.1	325	1	COBI_METJA
24	43	33.1	364	1	CA19_RAT
25	43	33.1	382	1	CA1_SVA0
26	43	33.1	382	1	FETB_HUMAN
27	43	33.1	390	1	COBI_MYCTU
28	43	33.1	481	1	GLGA_RHILTO
29	43	33.1	505	1	YTSJ_CAEEL
30	43	33.1	516	1	CAAD_PROME
31	43	33.1	628	1	HNFA_MOUSE
32	43	33.1	628	1	HNFA_RAT
33	43	33.1	921	1	HNFA_HUMAN
					CA19_HUMAN

34	43	33.1	921	1	CA19_MOUSE	Q05722 mus musculu
35	43	33.1	1133	1	RAD8_SCHPO	P36607 schizosacch
36	43	33.1	1137	1	PHYC_ORYSA	O92419 oryza sativ
37	42.5	32.7	193	1	VIF_SIVC2	P17284 chlamydiae
38	42.5	32.7	684	1	CA39_HUMAN	O14050 homo sapien
39	42.5	32.7	911	1	CPDI_MOUSE	P30999 mus musculu
40	42.5	32.7	1466	1	CA13_HUMAN	P02461 homo sapien
41	42.5	32.7	1669	1	CA14_HUMAN	P02462 homo sapien
42	42.5	32.7	1690	1	CA44_HUMAN	P53420 homo sapien
43	42	32.3	252	1	TFXE_RHILT	P42727 rhizobium 1
44	42	32.3	416	1	CINA_SYNY3	O55760 synechocyst
45	42	32.3	469	1	GLYC_YEAST	P37291 saccharomyc

## ALIGNMENTS

RESULT 1  
ID PURU\_HAEN STANDARD: PRT; 278 AA.  
AC Q03432;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)  
DE hydrolase).  
GN PURU OR H11588.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RS SEQUENCE FROM N.A.  
RP STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kesteven K., Sulten G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA McKenney K., Shiley R., Liu L.-I., Goddek A., Kelley J.M.,  
RA Scott J.D., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sudek D.M., Brandon R.C.,  
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
RA Grahn C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT Influenzae Rd.";  
RL Science 269:496-512(1995).  
RN [2]  
RP SEQUENCE OF 64-278 FROM N.A.  
RX STRAIN=RM 7004;  
RX MEDLINE=93328119; PubMed=8335255;  
RA Maskell D.J.;  
RT "Cloning and sequencing of the Haemophilus influenzae araC gene.";  
RL Gene 129:155-156(1993).  
CC -1- FUNCTION. PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES  
CC THE MAJOR SOURCE OF FORMATE FOR THE PRT-DEPENDENT SYNTHESIS OF  
CC 5-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC  
CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL  
CC (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +  
CC tetrahydrofolate.  
CC -1- ENZYME REGULATION: ACTIVATED BY METHIONINE, INHIBITED BY GLYCINE  
CC (BY SIMILARITY).  
CC -1- PATHWAY: DE NOVO PURINE BIOSYNTHESIS.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@sib.ch).
CC -----
DR EMBL; U32833; AAC23236.1; -
DR EMBL; L04686; AAA24942.1; -
DR PIR; P06066; P06066.
DR HSSP; P08179; IGRC.
DR TIGR; H11588; -
DR InterPro; IPR002912; ACT.
DR InterPro; IPR002376; formyl_transf.
DR Pfam; PF01842; ACT; 1.
KM Purine biosynthesis; Hydrolyase; One-carbon metabolism;
Complete proteome.
FT ACT_SITE 223 BY SIMILARITY.
FT CONFLICT 115 VIG -> RNR (IN REF. 2).
FT CONFLICT 138 HEN -> PK (IN REF. 2).
FT CONFLICT 205 K -> E (IN REF. 2).
SQ SEQUENCE 278 AA; 32173 MW; 7F375AB3C422EC4B CRC64;

Query Match 37.7%; Score 49; DB 1; Length 278;
Best Local Similarity 52.9%; Pred. No. 6.7;
Matches 9; Conservative 2; Mismatches 6; Gaps 0;

Oy 4 LPAVGLSPGEQEHYRG 20
Db 191 LPAFIGAKPYQQAAYKRG 207

RESULT 2
YAEFC_SCHPO STANDARD: PRT; 559 AA.
AC 009852:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative inorganic phosphate transporter C23D3.12.
GS SPMC23D3.12.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCB1_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Niblett D., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HIGH-AFFINITY TRANSPORTER FOR EXTERNAL INORGANIC
PHOSPHATE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRONG, TO
YEAST PHO84.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; Z64354; CAA91247.1; -
DR InterPro; IPR003662; sub-transporter.
DR Pfam; PF00083; sugar_tr.1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KM Hypothetical protein; Phosphate transport; Transmembrane.
FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 46 66 1 (POTENTIAL).
FT DOMAIN 67 94 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 95 115 2 (POTENTIAL).
FT DOMAIN 116 116 CYTOPLASMIC (POTENTIAL).

```

```

FT TRANSMEM 119 139 3 (POTENTIAL).
FT DOMAIN 140 144 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 145 165 4 (POTENTIAL).
FT DOMAIN 166 188 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 189 209 5 (POTENTIAL).
FT DOMAIN 210 224 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 225 245 6 (POTENTIAL).
FT DOMAIN 246 335 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 336 356 7 (POTENTIAL).
FT DOMAIN 357 381 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 382 402 8 (POTENTIAL).
FT DOMAIN 403 409 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 410 430 9 (POTENTIAL).
FT DOMAIN 431 438 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 439 459 10 (POTENTIAL).
FT DOMAIN 460 474 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 475 495 11 (POTENTIAL).
FT DOMAIN 496 497 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 498 518 12 (POTENTIAL).
FT DOMAIN 519 559 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 559 AA; 62316 MW; 2E8958F86C2092E2 CRC64;

Query Match 37.3%; Score 48.5; DB 1; Length 559;
Best Local Similarity 47.6%; Pred. No. 16;
Matches 10; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

Oy 8 VGLSPGEQEHY---RGVGV 25
Db 365 IGFSSGKNHYHTLMRGATGNL 365

RESULT 3
ARAB_YERPE STANDARD: PRT; 563 AA.
AC 0156543:
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE L-ribulokinase (EC 2.7.1.16).
DE ARAB OR YPO2254.
OS Versinia pestsis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCB1_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holtroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Versinia pestsis, the causative agent of plague.";
Nature 413:523-527(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-
phosphate.
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; AJ14151; CAC91038.1; ALT_INIT
DR Transferase; Kinase; Arabinose catabolism; Complete proteome.

```





RL J. Bacteriol. 177:1292-1298(1995).  
CC -1- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES  
CC THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF  
CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC  
CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.  
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +  
CC tetrahydrofolate.  
CC -1- ENZYME REGULATION: ACTIVATED BY METHIONINE, INHIBITED BY GLYCINE.  
CC -1- PATHWAY: DE NOVO PURINE BIOSYNTHESIS.  
CC -1- SUBUNIT: HOMOHXAMER.  
CC -1- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L20251; AAC36846.1; -;  
DR EMBL; M64673; AAA16860.1; ALT\_INIT.  
DR EMBL; AE000221; AAC74314.1; -;  
DR EMBL; D90758; BAA36100.1; -;  
DR EMBL; D90759; BAA36112.1; -;  
DR EMBL; D90852; BAA16026.1; -;  
DR PIR; C36871; C36871.  
DR HSSP; P08179; IGRC.  
DR Ecogene; EG11819; purU.  
DR InterPro; IPR002912; ACT.  
DR InterPro; IPR002376; formyl\_transf.  
DR Pfam; PF01842; ACT; 1.  
DR Pfam; PF00551; formyl\_transf; 1.  
KW Purine biosynthesis; Hydrolyase; One-carbon metabolism;  
KW Complete proteome.  
FT ACT\_SITE 225 BY SIMILARITY.  
SQ SEQUENCE 280 AA; 31934 MW; 5667406D2727A2C2 CRC64;  
  
Query Match 36.2%; Score 47; DB 1; Length 280;  
Best Local Similarity 52.9%; Pred. No. 13;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
  
OY 4 LPAVGLSPGEQYHRG 20  
|||:| | | | |  
Db 193 LPAFIGARPYHQAYERG 209  
  
RESULT 6  
PURU\_SHIFL STANDARD; PRT; 280 AA.  
AC P38480;  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4)  
DE hydrolyase).  
GN purU.  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=623;  
OX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-2A;  
RX MEDLINE=93023838; PubMed=1406252;  
RA Homocyst A.E., Tucker S.C., Maurelli A.T.;  
RT "Temperature regulation of Shigella virulence: Identification of the  
RT repressor gene virR, an analogue of hns, and partial complementation  
RT by tyrosyl transfer RNA (tRNA(Tyr)).";  
RL Mol. Microbiol. 6:2113-2124(1992).  
RN [2]  
RP IDENTIFICATION.

RX MEDLINE=94042872; PubMed=8226647;  
RA Nady P.L., McCorkle G., Zalkin H.;  
RT "purU, a source of formate for purt-dependent phosphoribosyl-N-  
RT formylglycinamide synthesis";  
RL J. Bacteriol. 175:7066-7073(1993).  
CC -1- FUNCTION: PRODUCES FORMATE FROM FORMYL-TETRAHYDROFOLATE. PROVIDES  
CC THE MAJOR SOURCE OF FORMATE FOR THE PURT-DEPENDENT SYNTHESIS OF  
CC 5'-PHOSPHORIBOSYL-N-FORMYLGLYCINAMIDE (FGAR) DURING AEROBIC  
CC GROWTH. HAS A ROLE IN REGULATING THE ONE-CARBON POOL.  
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + H(2)O = formate +  
CC tetrahydrofolate.  
CC -1- ENZYME REGULATION: ACTIVATED BY METHIONINE, INHIBITED BY GLYCINE  
CC (BY SIMILARITY).  
CC -1- PATHWAY: DE NOVO PURINE BIOSYNTHESIS.  
CC -1- SUBUNIT: HOMOHXAMER (BY SIMILARITY).  
CC -1- SIMILARITY: SOME, TO GAR TRANSFORMYLASE (PURN).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X66849; -; NOT\_ANNOTATED\_CDS.  
DR HSSP; P08179; IGRC.  
DR InterPro; IPR002912; ACT.  
DR InterPro; IPR002376; formyl\_transf.  
DR Pfam; PF01842; ACT; 1.  
DR Pfam; PF00551; formyl\_transf; 1.  
KW Purine biosynthesis; Hydrolyase; One-carbon metabolism.  
FT ACT\_SITE 225 BY SIMILARITY.  
SQ SEQUENCE 280 AA; 31877 MW; A60ACDD8979D4FB7 CRC64;  
  
Query Match 36.2%; Score 47; DB 1; Length 280;  
Best Local Similarity 52.9%; Pred. No. 13;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
  
OY 4 LPAVGLSPGEQYHRG 20  
|||:| | | | |  
Db 193 LPAFIGARPYHQAYERG 209  
  
RESULT 7  
PGCA\_CHICK STANDARD; PRT; 2109 AA.  
AC P07898; Q90991; Q90820; Q91047; Q90810;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core  
DE protein) (CSPCP).  
GN AGC1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
OX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WHITE LEKHORN; TISSUE=Embryo;  
RX MEDLINE=94043149; PubMed=8226878;  
RA Li H., Schwartz N.B., Vertel B.M.;  
RT "cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core  
RT protein and identification of a stop codon in the aggrecan gene  
RT associated with the chondrodys trophy, nanomelia.";  
RL J. Biol. Chem. 268:23504-23511(1993).  
RN [2]  
RP SEQUENCE OF 1042-1559 FROM N.A.  
RC TISSUE=Embryo;  
RX MEDLINE=90307744; PubMed=1694853;



```

FT DISULFID 1993 2009 BY SIMILARITY.
FT DISULFID 2024 2067 BY SIMILARITY.
FT DISULFID 2053 2080 BY SIMILARITY.
FT CARBOHYD 76 76 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 644 644 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 700 700 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 765 765 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1856 1892 MISSING (IN ISOFORM 2).
FT CONFLICT 362 362 E -> D (IN REF. 3).
FT CONFLICT 601 601 G -> D (IN REF. 3).
FT CONFLICT 1000 1000 P -> R (IN REF. 3).
FT CONFLICT 1029 1029 A -> P (IN REF. 3).
FT CONFLICT 1042 1043 VT -> PA (IN REF. 2).
FT CONFLICT 1251 1251 E -> D (IN REF. 2 AND 3).
FT CONFLICT 1587 1587 I -> T (IN REF. 5).
FT CONFLICT 1590 1590 I -> V (IN REF. 5).
FT CONFLICT 1594 1594 T -> S (IN REF. 5).
FT CONFLICT 1602 1610 IETSTVEI -> VLRCQSVLR (IN REF. 5).
FT CONFLICT 1603 1603 E -> A (IN REF. 3).
FT CONFLICT 1672 1672 S -> G (IN REF. 3).
FT CONFLICT 1796 1796 E -> G (IN REF. 3 AND 7).
FT CONFLICT 1988 1988 F -> S (IN REF. 6).
SQ SEQUENCE 2109 AA; 223492 MM; 7F824FD5B5A2ABDA CRC64;

```

```

Query Match 36.2%; Score 47; DB 1; Length 2109;
Best Local Similarity 47.6%; Pred. No. 1,le+02;
Matches 10; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

```

```

OY 4 LPAYVGLSPGEQEHRCGVV 24
DB 957 LVEVYTAAPGRQEGKSGIV 977

```

```

RESULT 8
COBI_MYCTU STANDARD; PRT; 508 AA.
ID COBI_MYCTU
AC Q10677;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cobalamin biosynthesis protein COBIJ [Includes: Precorrin-2 C20-
  methyltransferase (SPZMT); Precorrin-3 methylase (EC 2.1.1.-)].
DE COBIJ OR RV2066 OR MT2126 OR MTCY49.05.
GN COBIJ OR RV2066 OR MT2126 OR MTCY49.05.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
  Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
  Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekela F.,
  Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
  Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
  Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
  Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
  Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
  Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
  complete genome sequence."
RT Nature 393:537-544 (1998).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
  Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,

```

```

RA Kolonay J.F., Nelson W.C., Umamay L.A., Ermolaeva M.D., Salzberg S.L.,
  Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
  Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
  laboratory strains."
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: METHYLASES PRECORIN-2 AT THE C-20 POSITION TO PRODUCE
  PRECORIN-3A (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 = S-
  adenosyl-L-homocysteine + precorrin-3a.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUWT, CYSG, CBI/COBM
  AND CBI/COBI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73966; CAA98214.1; -
DR EMBL; AE007063; AAK46406.1; -
DR TIGR; MT2126; -
DR TIGR; RV2066; -
DR InterPro; IPR000878; Corrin_porph_methyltrnf.
DR InterPro; IPR003043; Uropor_methyltrnsf.
DR Pfam; PF00590; TP_methylase; 2.
DR PROSITE; PS00839; SUWT_1; 1.
DR PROSITE; PS00840; SUWT_2; 1.
DR Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
  Methyltransferase; Multifunctional enzyme; Complete proteome.
KW DOMAIN 1 243 PRECORIN-2 C20-METHYLTRANSFERASE.
FT DOMAIN 244 508 PRECORIN-3 METHYLASE.
SQ SEQUENCE 508 AA; 53910 MM; 95AC06F02C4DC1 CRC64;

```

```

Query Match 35.4%; Score 46; DB 1; Length 508;
Best Local Similarity 47.1%; Pred. No. 34;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

```

OY 1 MAGLPAYVGLSPGEQEV 17
DB 245 LTGVYAVVGLPGDSBW 261

```

```

RESULT 9
PTSB_VIBAL STANDARD; PRT; 479 AA.
ID PTSB_VIBAL
AC P22825;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-
  permease IIBC component) (Phosphotransferase enzyme II, BC component)
  (EC 2.7.1.69) (EII-SCR).
DE SCRA.
GN Vibrio alginolyticus.
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=663;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91071601; PubMed=2174811;
RA Blatch G.L., Scholle R.R., Woods D.R.;
RT "Nucleotide sequence and analysis of the Vibrio alginolyticus sucrose
  uptake-encoding region."
RL Gene 95:17-23 (1990).
RN [2]
RP SEQUENCE OF 1-12 FROM N.A.
RX MEDLINE=91285433; PubMed=2060795;
RA Blatch G.L., Woods D.R.;
RT "Nucleotide sequence and analysis of the Vibrio alginolyticus scr

```

```

RT      retroviral-encoding gene (scrv).".
RL      Gene.101:45-50(1991)".
CC      -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC      SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC      TRANSPORT SYSTEM. THE IIIC DOMAIN CONTAINS THE SUGAR BINDING SITE
CC      AND THE TRANSMEMBRANE CHANNEL. THE IIA DOMAIN CONTAINS THE PRIMARY
CC      PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIB TRANSFERS ITS
CC      PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC      THE SUGAR.
CC      -1- CATALYTIC ACTIVITY: protein N-phosphohistidine + sugar = protein
CC      histidine + sugar phosphate.
CC      -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC      -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC      -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on
CC      its use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@sdb-sib.ch).
CC      -----
CC      EMBL: M76768; AAA27555.1; -.
CC      EMBL: M35009; AAA27557.2; -.
CC      PIR: JQ0781; JQ0781.
CC      HSSP: P05053; IIBA.
CC      InterPro: IPRO01996; PTS_EIIB.
CC      InterPro: IPRO03352; PTS_EIIC.
CC      Pfam: PF00367; PTS_EIIB; 1.
CC      Pfam: PF02378; PTS_EIIC; 1.
CC      ProDom: PD001476; PTS_EIIB; 1.
CC      PROSITE: PS01035; PTS_EIIB_CYS_1.
CC      Phosphotransferase system: Sugar transport: Transferase;
CC      Transmembrane; Inner membrane; Phosphorylation.
CC      DOMAIN          ?           ?           EIIB DOMAIN.
CC      FT              ?           479        EIIC DOMAIN.
CC      MOD_RES         26          26         PHOSPHORYLATION (BY SIMILARITY).
CC      MOD_RES         324         324        PHOSPHORYLATION (BY SIMILARITY).
CC      SEQUENCE       479 AA; 49890 MW; FBF906B5170E3EB7 CMC64;
CC      -----
Query Match               34.6%; Score 45; DB 1; Length 479;
Best Local Similarity    33.3%; Pred. NO. 45;
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
QY      2 AGPAVWGLSGEQEEYHRCGV 22
DB      431 AGPGIIISITPDKIGYTAGM 451
||| : : | : | : | :
ACON_MYCAV STANDARD; PRT; 961 AA.
AC      008451;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)
DE      Aconitate hydratase (EC 4.2.1.3) (citrate hydro-lyase) (Aconitase).
GN      ACN.
OS      Mycobacterium avium.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1764;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=GIRIO;
RX      MEDLINE=98195739; PubMed=9534249;
RA      Labo M., Guibert L., de Rossi E., Speziale P., Ricciardi G.;
RT      "Determination of a 15437 bp nucleotide sequence around the inhA gene
RT      of Mycobacterium avium and similarity analysis of the products of
RT      putative ORFs.";
Microbiology 144:807-814(1998).
```

```

CC -1- CATALYTIC ACTIVITY: citrate - cis-aconitate + H(2)O.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- MISCELLANEOUS: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE (3FE-4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE CATALYTIC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND ISOCITRATE.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/TPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF02133; AAC46192.1; -.
CC InterPro: IPR001030; Aconitase.
CC InterPro: IPR000573; Aconitase_C.
CC Pfam: PF00330; aconitase; 2.
CC Pfam: PF00694; Aconitase_C; 1.
CC Prodom: PD000511; Aconitase; 2.
CC PROSITE: PS00450; ACONITASE; 1.
CC PROSITE: PS01244; ACONITASE; 2; 1.
CC Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S.
CC METAL 499 499 IRON (IRON-SULFUR CLUSTER)
CC FT METAL (BY SIMILARITY).
CC FT METAL 565 565 IRON (IRON-SULFUR CLUSTER)
CC FT METAL (BY SIMILARITY).
CC FT METAL 568 568 IRON (IRON-SULFUR CLUSTER)
CC FT METAL (BY SIMILARITY).
CC FT METAL (BY SIMILARITY).
CC SQ SEQUENCE 961 AA; 104025 MW; 689E66D85E31F596 CRC64;

Query Match 34.6%; Score 45; DB 1; Length 961;
Best Local Similarity 48.0%; Pred. No. 93;
Matches 12; Conservative 5; Mismatches 6; Indels 2; Gaps 2;

Qy 2 AGLPAAVGL-SFGEQEHARGGVVL 25
| ||| : ||| : ||| : ||| :
Db 927 AEFDAVVRIDTPGEADYYRNG-GIL 950

RESULT 11
SODC_XIPGL STANDARD; PRT; 151 AA.
AC P03946;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).
OS Xiphias gladius (Swordfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Scombroidei;
OC Xiphiidae; Xiphias.
OX NCBI_Taxid=8245;
OC [1]
RN
RP SEQUENCE.
RX MEDLINE=85076642; PubMed=6510412;
RA Rocha H.A., Bannister W.H., Bannister J.V.;
RT "The amino-acid sequence of copper/zinc superoxide dismutase from
RT swordfish liver. Comparison of copper/zinc superoxide dismutase
RT sequences."
RL Eur. J. Biochem. 145:477-484(1984).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -1- COFACTOR: Copper and zinc.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.

```

DR PIR: A00516; DSMFCZ.  
 DR HSSP: P00442; 1CBZ.  
 DR InterPro: IPR001424; SOD\_CU\_ZN.  
 DR Pfam: PF000080; sodcu.1  
 DR PRINTS: PR000080; CUZNDISMTASE.  
 DR Prodom: PD000469; SOD\_CU\_ZN.1.  
 DR PROSITE: PS00087; SOD\_CU\_ZN.1;  
 DR PROSITE: PS00332; SOD\_CU\_ZN.2; 1.  
 KW Oxidoreductase; Copper; Zinc.  
 FT METAL 46 46 COPPER.  
 FT METAL 48 48 COPPER.  
 FT METAL 63 63 COPPER AND ZINC.  
 FT METAL 71 71 ZINC.  
 FT METAL 80 80 ZINC.  
 FT METAL 83 83 ZINC.  
 FT METAL 119 119 COPPER.  
 FT DISULFID 57 145  
 FT ACT\_SITE 142 142  
 FT SEQUENCE 151 AA; 15517 MW; E7FA3EF2F934E574 CRC64;  
 SQ  
 Query Match 33.8%; Score 44; DB 1; Length 151;  
 Best Local Similarity 53.3%; Pred. No. 19;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 9 GLSPGEQYHRCVG 23  
 Db 37 GLTPEGHGFVHGF 51

RESULT 12  
 EMOB\_RABIT STANDARD: PRT: 433 AA.

AC P25704; Q9N0N6; (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Beta enolase (EC 4.2.1.11) (2-phospho-D-glycerate hydro-lyase)  
 DE (Skeletal muscle enolase).  
 GN ENO3.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=91113295; PubMed=2275753;  
 RA "Chn C.C.Q.;  
 RT "The primary structure of rabbit muscle enolase.";  
 RL J. Protein Chem. 9:427-432(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Zheng S.-X.;  
 RT "The cdna cloning of rabbit muscle-specific enolase gene, site  
 RT directed mutagenesis (E417L) of the gene, expression of the wild-type  
 RT and mutant genes in *Escherichia coli*.";  
 RL Thesis (1995), Concordia University, Montreal / Quebec, Canada.  
 RN [3]  
 RP REVISIONS TO 296-308 AND 314.  
 RA Kornblatt M.J.; Zheng S.-X.; Lamande N.; Lazar M.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: 2-phospho-D-glycerate = phosphoenolpyruvate +  
 CC H(2O).  
 CC -1- COFACTOR: MAGNESIUM IS REQUIRED FOR CATALYSIS AND FOR STABILIZING  
 CC THE DIMER.  
 CC -1- PATHWAY: GLYCOLYSIS.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: THREE CLASSES OF ENOLASE ISOENZYMES HAVE BEEN  
 CC IDENTIFIED IN MAMMALIAN TISSUES. ISOENZYME ALPHA IS PRESENT IN  
 CC MOST TISSUES, BETA IS LOCALIZED IN MUSCLE TISSUE, AND GAMMA IS  
 CC FOUND ONLY IN NERVOUS TISSUE.

CC -1- SIMILARITY: BELONGS TO THE ENOLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF260259; AAF71925.2; -  
 CC PIR: A37210; A37210.  
 CC HSSP: P56252; 1PDZ.  
 CC InterPro: IPR000941; Enolase.  
 CC Pfam: PF00113; enolase.1.  
 CC PRINTS: PR00148; ENOLASE.  
 CC Prodom: PD000902; ENOLASE.  
 CC DR PROSITE: PS00164; ENOLASE.1.  
 CC DR PROSITE: PS00164; ENOLASE.1.  
 KW Lyase; Glycolysis; Magnesium; Multigene family.  
 FT INIT\_MET 0 0  
 FT ACT\_SITE 157 157  
 FT METAL 244 244 BY SIMILARITY.  
 FT METAL 292 292 MAGNESIUM (BY SIMILARITY).  
 FT METAL 317 317 MAGNESIUM (BY SIMILARITY).  
 FT METAL 317 317 N -> D (IN REF. 1).  
 FT CONFLICT 253 253 DMATWTSFLS -> GDMGAWGRFLA (IN REF. 1).  
 FT CONFLICT 298 308  
 FT CONFLICT 314 314 I -> V (IN REF. 1)  
 FT SEQUENCE 433 AA; 46937 MW; F2348C53E508DF04 CRC64;  
 SQ

Query Match 33.8%; Score 44; DB 1; Length 433;  
 Best Local Similarity 40.9%; Pred. No. 56;  
 Matches 9; Conservative 3; Mismatches 8; Indels 2; Gaps 1;

Qy 2 AGLP-AVGLSPGEQYHRCG 21  
 Db 233 AGYPDKVYIGMDVAASEFHRNG 254

RESULT 13  
 ARAB\_ECO57 STANDARD: PRT: 565 AA.  
 ID ARAB\_ECO57  
 AC P58541;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE L-ribulokinase (EC 2.7.1.16).  
 DE ARAB OR 20072 OR ECS0067.  
 OS *Escherichia coli* O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Escherichia*.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T.; Plunkett G. III; Burland V.; Mau B.; Glasner J.D.;  
 RA Rose D.J.; Mayhew G.F.; Evans P.S.; Gregor J.; Kirkpatrick H.A.;  
 RA Postal G.; Hackett J.; Klink S.; Boutin A.; Shao Y.; Miller L.;  
 RA Grobeck E.J.; Davis N.W.; Lim A.; Dimalanta E.T.; Potamoustis K.;  
 RA Apodaca J.; Anantharaman T.S.; Lin J.; Yen G.; Schwartz D.C.;  
 RA Welch R.A.; Blattner F.R.;  
 RA "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RT Nature 409:529-533(2001).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T.; Makino K.; Ohnishi M.; Kurokawa K.; Ishii K.; Yokoyama K.;  
 RA Han C.-G.; Ohnishi E.; Nakayama K.; Murata T.; Tanaka M.; Tobe T.;  
 RA Tada T.; Takami H.; Honda T.; Sasaki K.; Ogasawara N.; Yasunaga T.;  
 RA Kihara S.; Shiba T.; Hattori M.; Shinagawa H.;  
 RA "Complete genome sequence of enterohaemorrhagic *Escherichia coli*

```

RT 0157:47 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose -> ADP + L-ribulose 5-
CC phosphate.
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC -1- SIMILARITY: BELONGS TO THE RIBULOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A6005183; AAC54367.1; -
DR EMBL: AP002550; BAB3490.1; -
KM Transferase: Kinase; Arabinose catabolism.
FT INT_MET 0 BY SIMILARITY.
SQ SEQUENCE 565 AA; 61096 MW; 45E7D9AA096B9CAD CRC64;

Query Match 33.8%; Score 44; DB 1; Length 565;
Best Local Similarity 47.6%; Pred. No. 74;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 3 GLPAVGLSPGEQHYHGCVG 23
    ||| | | | | | | | | |
Db 260 GLPESVVISGGAFCDHMGAVG 280

RESULT 14
ARAB_ECOLI
ID ARAB_ECOLI STANDARD; PRT; 565 AA.
AC P08204; P78041.
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-MAR-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE L-ribulokinase (EC 2.7.1.16).
GN ARAB OR B0063.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87163495; PubMed=3549454;
RA Lee N., Glelow W., Martin R., Hamilton E., Fowler A.;
RT "The organization of the arabid operon of Escherichia coli.";
RL Gene 47:231-244(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori I., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizubuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 1-13 FROM N.A., AND SEQUENCE OF 1-13.
RX MEDLINE=77102763; PubMed=189315;
RA Lee N., Carbon J.;

```

```

RT "Nucleotide sequence of the 5' end of arabid operon messenger RNA in
RT Escherichia coli B/r.";
RL Proc. Natl. Acad. Sci. U.S.A. 74:49-53(1977).
RN [5]
RP SEQUENCE OF 1-5 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=79005683; PubMed=357433;
RA Smith B.R., Schleif R.;
RT "Nucleotide sequence of the L-arabinose regulatory region of
RT Escherichia coli K12.";
RL J. Biol. Chem. 253:6931-6933(1978).
CC -1- CATALYTIC ACTIVITY: ATP + L-ribulose -> ADP + L-ribulose 5-
CC phosphate.
CC -1- PATHWAY: L-arabinose catabolism; second step.
CC -1- SIMILARITY: BELONGS TO THE RIBULOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15263; AAA23462.1; -
DR EMBL: J01641; AAA23467.1; -
DR EMBL: AE000116; AAC73174.1; -
DR EMBL: K01304; AAA23465.1; -
DR EMBL: D10483; BAA01334.1; -
DR PIR: B29022; KIBCRU.
DR PIR: S40579; S40579.
DR ECODBASE: D055.0; 6TH EDITION.
DR Ecogene: EG10053; arab.
DR InterPro: IPR000577; FCGY_kin.
DR Pfam: PF00370; FCGY_2.
DR Pfam: PF02782; FCGY_C_1.
KM Transferase: Kinase; Arabinose catabolism; Complete proteome.
FT INT_MET 0
FT CONFLICT 22 22 T -> S (IN REF. 1 AND 2).
FT CONFLICT 126 127 EA -> RS (IN REF. 1 AND 2).
FT CONFLICT 349 349 G -> S (IN REF. 1 AND 2).
FT CONFLICT 364 364 T -> A (IN REF. 1 AND 2).
FT CONFLICT 402 402 T -> S (IN REF. 1 AND 2).
FT CONFLICT 524 524 C -> R (IN REF. 1 AND 2).
SQ SEQUENCE 565 AA; 60958 MW; DB9B39AD92062484 CRC64;

Query Match 33.8%; Score 44; DB 1; Length 565;
Best Local Similarity 47.6%; Pred. No. 74;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 3 GLPAVGLSPGEQHYHGCVG 23
    ||| | | | | | | | | |
Db 260 GLPESVVISGGAFCDHMGAVG 280

RESULT 15
ARAB_SALTI
ID ARAB_SALTI STANDARD; PRT; 568 AA.
AC P58342;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE L-ribulokinase (EC 2.7.1.16).
GN ARAB OR STY0120.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CT18;
RX MEDLINE=21534947; PubMed=11677608;

```

RA Parthill J., Dungan G., James K.D., Thomson N.R., Pickard D., Main J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebathia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*  
 RT enterica serovar Typhim CT18.";  
 RT Nature 413:848-852(2001).  
 CC -1- CATABOLIC ACTIVITY: ATP + L-ribulose = ADP + L-ribulose 5-  
 CC phosphate.  
 CC -1- PATHWAY: L-arabinose catabolism: second step.  
 CC -1- SIMILARITY: BELONGS TO THE RIBULOXINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL627265; CAD01260.1;  
 KW transferase; kinase; Arabinose catabolism; Complete proteome.  
 FT INIT\_MET 0 BY SIMILARITY  
 SQ SEQUENCE 568 AA; 61643 MW; A1515DB7249A0F75 CRC64;

Query Match 33.8%; Score 44; DB 1; Length 568;  
 Best Local Similarity 47.6%; Pred. No. 75;  
 Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Oy 3 GLPAVVGSLPGEQYHRGCVG 23  
 ||| | : | | : | | |  
 Db 260 GLPESVVISGAFDCMGCAVG 280

Search completed: September 3, 2002, 15:25:32  
 Job time: 200 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:19:27 : Search time 14.82 Seconds  
(without alignments)  
162.094 Million cell updates/sec

Title: US-09-786-214-5

Sequence: 130  
1 MACLPVAVGLSPGEYHRGVGV 25

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR-71:\*

1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	39.2	426	2 D72635	hypothetical prote
2	50.5	38.8	615	2 AC3554	galactonate dehydr
3	50	38.5	277	2 F82130	formyltetrahydrofo
4	50	38.5	381	2 F75270	cytochrome P450 -
5	49.5	38.1	603	2 H98330	hypothetical prote
6	49.5	38.1	603	2 AE2952	dihydroxy-acid deh
7	49	37.7	118	2 S38491	lg heavy chain - h
8	49	37.7	278	2 E64131	formyltetrahydrofo
9	48.5	37.3	559	2 S62503	inorganic phosphat
10	48	36.9	522	2 T44369	pyruvate, water dik
11	48	36.9	567	2 AF0274	ribulokinase (EC 2
12	48	36.9	780	2 D75361	pentolillin binding
13	48	36.9	792	2 B82752	formyltetrahydrofo
14	47	36.2	280	2 C36871	formyltetrahydrofo
15	47	36.2	280	2 F90845	formyltetrahydrofo
16	47	36.2	280	2 E85703	hypothetical prote
17	47	36.2	280	2 AF0649	formyltetrahydrofo
18	47	36.2	293	2 E71819	formyltetrahydrofo
19	47	36.2	293	2 B84659	formyltetrahydrofo
20	47	36.2	661	2 G84511	hypothetical prote
21	47	36.2	2109	1 T50421	aggreacan precursor
22	46.5	35.8	178	2 G82977	hypothetical prote
23	46.5	35.8	553	2 T36580	hypothetical prote
24	46.5	35.8	656	2 T36580	hypothetical prote
25	46	35.4	192	2 B81692	pentolillin-binding
26	46	35.4	271	2 AB2545	hypothetical prote
27	46	35.4	274	2 C82841	cysteine proteinas
28	46	35.4	287	2 F81350	formyltetrahydrofo
29	46	35.4	300	2 B70871	hypothetical prote
				2 C69857	formyltetrahydrofo

30	46	35.4	508	2 E70764	probable cobi prot
31	46	35.4	638	2 F75547	anthranilate synth
32	46	35.4	1196	2 H87542	hydantoinase/oxopr
33	45.5	35.0	717	2 T50685	delital-pyrroline-5
34	45	34.6	104	2 C72637	hypothetical prote
35	45	34.6	158	2 E97500	hypothetical prote
36	45	34.6	158	2 A12718	conserved hypotet
37	45	34.6	211	2 AE3572	carbonic anhydrase
38	45	34.6	217	2 E72047	conserved hypotet
39	45	34.6	217	2 D86576	transcription regu
40	45	34.6	340	2 A12075	hypothetical prote
41	45	34.6	463	2 T49460	hypothetical prote
42	45	34.6	479	2 F82432	PTS system, sucros
43	45	34.6	479	2 F82432	sucrose uptake pro
44	45	34.6	582	2 T34538	hypothetical prote
45	45	34.6	940	2 JE0291	FB19 protein - hum

#### ALIGNMENTS

```

RESULT 1
D72635
hypothetical protein APE1539 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence=revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: D72635
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A:Reference number: A72450; MUID:99310339
A:Accession: D72635
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-426 <KAW>
A:Cross-references: DDBJ:AP000061; NID:g5104821; PIDN:BA80538.1; PID:d1044324; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1539
C:Superfamily: Aeropyrum pernix hypothetical protein APE1539

Query Match          39.2%   Score 51; DB 2; Length 426;
Best Local Similarity 45.5%   Pred. No. 15;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY      3 GLPAVGLSPGEYHRGVGV 24
Db      278 GLADIDIEPRGRERREGVRV 299

RESULT 2
AC3554
galactonate dehydratase (EC 4.2.1.6) [Imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence=revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AC3554
R:Deivechilo, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Heglus, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD2522; PMID:11756688
A:Accession: AC3554
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-615 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL53598.1; PID:g17984511; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME10356
A:Map position: II
C:Superfamily: dihydroxy-acid dehydratase

```

C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 38.8%; Score 50.5; DB 2; Length 615;  
Best Local Similarity 45.8%; Pred. No. 26;  
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

OY 4 LPAYVGLSP-----GEOEYHRCGV 22  
DB 324 IPLVNLQPAGEYLGEDYTHAGCV 347

RESULT 3

F82130  
Formyltetrahydrofolate deformylase VC1992 [imported] - Vibrio cholerae (strain N16961 se  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: F82130  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.  
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 405, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: F82130  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-277 <HEI>  
A:Cross-references: GB:AE004274; GB:AE003852; NID:q9656533; PIDN:AAF95140.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC1992  
A:Map position: 1  
C:Superfamily: phosphoribosylglycinamide formyltransferase; phosphoribosylglycinamide fo

Query Match 38.5%; Score 50; DB 2; Length 277;  
Best Local Similarity 52.9%; Pred. No. 13;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 4 LPAYVGLSPGQEHYHRC 20  
DB 190 LPATGAKPYOQAYERG 206

RESULT 4

F75270  
Cytochrome P450 - Deinoxoccus radiodurans (strain R1)  
C:Species: Deinoxoccus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: F75270  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M., Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinoxoccus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: F75270  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-381 <WHI>  
A:Cross-references: GB:AE002076; GB:AE000513; NID:q6460285; PIDN:AAI12016.1; PID:q646029  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR2473  
A:Map position: 1

Query Match 38.5%; Score 50; DB 2; Length 381;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GLPAYVGLSP 12

DB 51 GLPAYVGLSP 60

RESULT 5

H98330  
Hypothetical protein AGR\_L3190 [imported] - Agrobacterium tumefaciens (strain C58, C  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: H98330  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Matkeitz,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
A:Reference number: A97359; PMID:11743194  
A:Accession: H98330  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-603 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK90170.1; PID:q15160173; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L3190  
A:Map position: linear chromosome  
C:Superfamily: dihydroxy-acid dehydratase

Query Match 38.1%; Score 49.5; DB 2; Length 603;  
Best Local Similarity 45.8%; Pred. No. 36;  
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

OY 4 LPAYVGLSP-----GEOEYHRCGV 22  
DB 312 VPLVNLQPAGEYLGEDYTHAGCV 335

RESULT 6

AE2952  
dihydroxy-acid dehydratase [imported] - Agrobacterium tumefaciens (strain C58, Dupont  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
C:Accession: AE2952  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AE2952  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-603 <KUR>  
A:Cross-references: GB:AE006689; PIDN:AAI44035.1; PID:q17741597; GSPDB:GN00167  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: hlyD  
A:Map position: linear chromosome  
C:Superfamily: dihydroxy-acid dehydratase

Query Match 38.1%; Score 49.5; DB 2; Length 603;  
Best Local Similarity 45.8%; Pred. No. 36;  
Matches 11; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

OY 4 LPAYVGLSP-----GEOEYHRCGV 22  
DB 312 VPLVNLQPAGEYLGEDYTHAGCV 335

RESULT 7  
S38491  
Ig heavy chain - human (fragment)

C:Species: Homo sapiens (man)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999  
C:Accession: S38491  
R:Authors: J.D.; Ouehand, W.H.; Bye, J.M.; Finern, R.; Gorlick, B.D.; Voak, D.; Thorpe, S.  
submitted to the EMBL Data Library, June 1993  
A:Description: Human antibody fragments specific for human blood group antigens from a B  
A:Reference number: S38488  
A:Accession: S38491  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-118 <MAR>  
A:Cross-references: EMBL:Z23032; NID:G414029; PIDN:CAA80567.1; PID:G414030  
C:Superfamily: Immunoglobulin V region; Immunoglobulin homology  
C:Keywords: heterotrimer; Immunoglobulin  
F:15-97/Domain: Immunoglobulin homology <IMM>

Query Match 37.7%; Score 49; DB 2; Length 118;  
Best Local Similarity 50.0%; Pred. No. 7.4;  
Matches 10; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

OY 3 GLPVGSLSPGEQYHRCV 22  
DB 44 GLEWVGIGPGGTYHRCV 63

RESULT 8  
E64131  
formyltetrahydrofolate deformylase (EC 3.5.1.10) - Haemophilus influenzae (strain Rd KW2  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 17-Mar-2000  
C:Accession: E64131; PNO606  
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: E64131  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-278 <TRGR>  
A:Cross-references: GB:U32833; GB:LA2023; NID:G1574432; PIDN:AC33236.1; PID:G1574433; T  
R:Maskell, D.  
Gene 129, 155-156, 1993  
A:Title: Cloning and sequencing of the Haemophilus influenzae arca gene.  
A:Reference number: JN0758; MUID:93328119  
A:Accession: PNO606  
A:Molecule type: DNA  
A:Residues: 64-114 'RNR', 118-137 'PK', 141-204 'E', 206-278 <MAS>  
A:Cross-references: GB:U04686; NID:G148863; PIDN:AAA24942.1; PID:G148864  
C:Genetics:  
A:Gene: purN  
C:Function:  
A:Description: catalyzes hydrolysis of 10-formyltetrahydrofolate to tetrahydrofolate and  
A:Pathway: one-carbon metabolism  
A>Note: activated by methionine and inhibited by glycine  
C:Superfamily: phosphoribosylglycinamide formyltransferase; phosphoribosylglycinamide fo  
F:85-277/Domain: phosphoribosylglycinamide formyltransferase homology <PRGF>  
F:223/Active site: Asp #status predicted

Query Match 37.7%; Score 49; DB 2; Length 278;  
Best Local Similarity 52.9%; Pred. No. 19;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 4 LPVVGSLSPGEQYHRCV 20  
DB 191 LPATGAKPTQDAVKRG 207

RESULT 9  
S62503  
Inorganic phosphate transporter - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 10-Dec-1999  
C:Accession: T38287; S62503  
R:Niblett, D.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: 221783  
A:Accession: T38287  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-559 <NT2>  
A:Cross-references: EMBL:Z64354; NID:G1039338; PIDN:CAA91247.1; PID:G1039350; GSPDB:G  
A:Experimental source: strain 972h-; cosmid c23D3  
C:Genetics:  
A:Gene: SPDB:SPAC23D3.12  
A:Map position: 1R  
C:Superfamily: probable inorganic phosphate transport protein PHO84

Query Match 37.3%; Score 48.5; DB 2; Length 559;  
Best Local Similarity 47.6%; Pred. No. 46;  
Matches 10; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

OY 8 VGLSPGEQYH---RGVGV 25  
DB 365 IGFSSGKNEYHTLMGAI 385

RESULT 10  
T44369  
pyruvate, water dikinase (EC 2.7.9.2) [imported] - Deinococcus radiodurans (fragment)  
C:Species: Deinococcus radiodurans  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 02-Sep-2000  
C:Accession: T44369  
R:Narumi, T.; Islam, S.; Cherden, K.; Kikuchi, M.; Watanabe, H.; Kitayama, S.; Yamam  
submitted to the EMBL Data Library, August 1998  
A:Description: I58301: the second insertion sequence element from Deinococcus radiodu  
A:Reference number: 222755  
A:Accession: T44369  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-522 <NAR>  
A:Cross-references: EMBL:AB016803; PIDN:BAA32387.1  
A:Experimental source: strain KD8301  
C:Genetics:  
A>Note: ppsA  
C:Superfamily: Escherichia coli pyruvate, water dikinase; phosphotransferase system en  
C:Keywords: transferase

Query Match 36.9%; Score 48; DB 2; Length 522;  
Best Local Similarity 50.0%; Pred. No. 51;  
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 3 GLPVGSLSPGEQYHRCV 20  
DB 160 GIPAVGTGATRELHNG 177

RESULT 11  
AF0274  
ribulokinase (EC 2.7.1.16) [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AF0274  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M  
deno-Tarrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G  
11, M.; Rutherford, K.; Skellern, M.; Skellern, J.; Stevens, K.; Whitehead, S.; Barrell  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360



Query Match 36.2% Score 47; DB 2; Length 280;  
Best Local Similarity 52.9% Pred. No. 36;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 LPAVVGSLPGEQYHRG 20  
|||:| | | | |  
Db 193 LPAFIGARPYHQAYERG 209

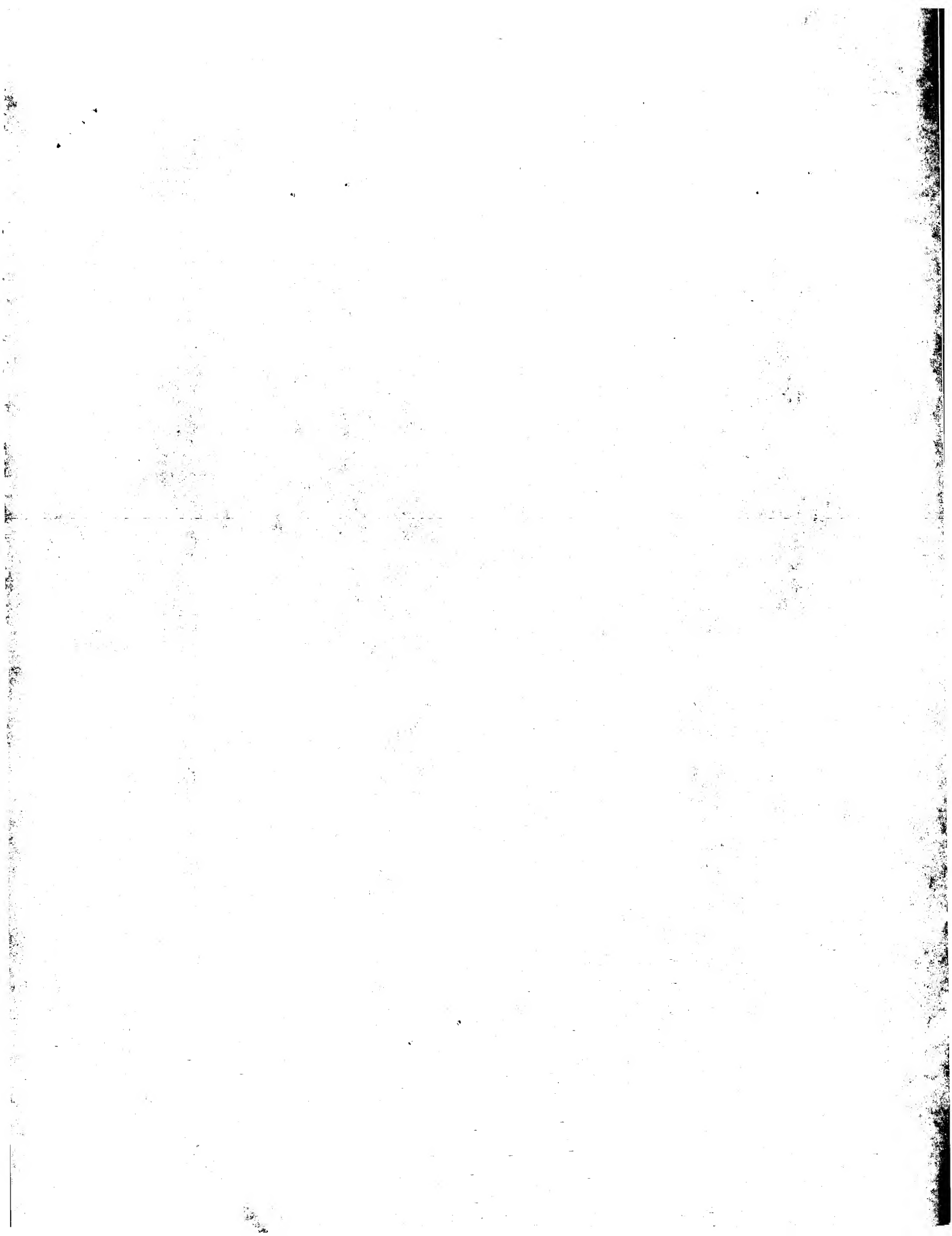
RESULT 15

F90845  
Formyltetrahydrofolate deformylase [Imported] - Escherichia coli (strain O157:H7, substr  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: F90845  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: F90845  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-280 <HA>  
A:Cross-references: GB:BA000007; PIDN:BA035157.1; PID:q13361199; GSPDB:GM00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECS1734  
C:Superfamily: phosphoribosylglycinamide formyltransferase; phosphoribosylglycinamide fo

Query Match 36.2% Score 47; DB 2; Length 280;  
Best Local Similarity 52.9% Pred. No. 36;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 LPAVVGSLPGEQYHRG 20  
|||:| | | | |  
Db 193 LPAFIGARPYHQAYERG 209

Search completed: September 3, 2002, 15:22:32  
Job time: 185 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:15:52 : Search time 30.08 Seconds

(without alignments)  
92.315 Million cell updates/sec

Title: US-09-786-214-5

Perfect score: 130  
Sequence: 1 MAGLPAYVGLSPGEYHRGCVYL 25

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq\_032802:\*  
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130	100.0	25	21	AA1984264
2	106	81.5	20	21	AA1984265
3	75	57.7	15	21	AA1984269
4	72	55.4	14	21	AA1984266
5	68	52.3	13	21	AA1984267
6	65	50.0	13	21	AA1984268
7	53	40.8	394	22	AAU43816
8	51.5	39.6	191	21	AA1984269
9	48.5	38.1	192	21	AA1984270
10	48.5	37.3	192	21	AA1984271
11	47.5	36.5	114	22	AA1984272

12	47	36.2	293	19	AA1984285	H. pylori GHP 997
13	46	35.4	46	21	AA158745	Somatosatatin anlig
14	46	35.4	234	22	AA158745	Human immune syste
15	46	35.4	300	20	AA158745	A formate transpor
16	46	35.4	921	20	AA158745	Human collagen IX
17	46	35.4	1121	22	AA158745	Novel human diago
18	45.5	35.0	280	22	AA158745	Novel human diago
19	45.5	35.0	348	21	AA158745	Wheat branched cha
20	45.5	35.0	853	22	AA158745	Novel human diago
21	45	34.6	116	22	AA158745	Novel human diago
22	45	34.6	207	22	AA158745	Novel human diago
23	45	34.6	238	20	AA158745	Novel human diago
24	45	34.6	253	22	AA158745	Chlamydia pneumoni
25	45	34.6	292	22	AA158745	Amino acid sequenc
26	45	34.6	328	21	AA158745	Amino acid sequenc
27	45	34.6	334	22	AA158745	Human membrane-ass
28	45	34.6	334	22	AA158745	Human guanine exch
29	45	34.6	338	22	AA158745	Human protein SEQ
30	45	34.6	342	21	AA158745	Amino acid sequenc
31	45	34.6	353	22	AA158745	Human prostate can
32	45	34.6	361	22	AA158745	Amino acid sequenc
33	45	34.6	399	22	AA158745	Human colon cancer
34	45	34.6	403	22	AA158745	Amino acid sequenc
35	45	34.6	438	22	AA158745	Human full-length
36	45	34.6	444	21	AA158745	Amino acid sequenc
37	45	34.6	456	22	AA158745	Amino acid sequenc
38	45	34.6	456	22	AA158745	Human protein SEQ
39	45	34.6	484	22	AA158745	Human polypeptide
40	45	34.6	484	22	AA158745	Amino acid sequenc
41	45	34.6	532	22	AA158745	Amino acid sequenc
42	45	34.6	547	22	AA158745	Amino acid sequenc
43	45	34.6	578	22	AA158745	Amino acid sequenc
44	45	34.6	593	22	AA158745	Amino acid sequenc
45	45	34.6	678	22	AA158745	Amino acid sequenc

#### ALIGNMENTS

RESULT 1  
ID AAY84264 standard; Peptide: 25 AA.  
AC AAY84264;  
XX  
XX  
DT 12-JUL-2000 (first entry)  
XX  
XX  
DE Peptide of alternate reading frame of macrophage colony stimulating gene.  
KW Renal cell carcinoma: antigen; cytotoxic T lymphocyte;  
KW tumour rejection antigen; macrophage colony stimulating gene;  
KW macrophage-colony stimulating factor; antigen presenting cell;  
KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.  
OS Homo sapiens.  
PN WO200013699-A1.  
XX  
XX  
PD 16-MAR-2000.  
XX  
XX  
PF 03-SEP-1999; 99WO-US20344.  
XX  
XX  
PR 04-SEP-1998; 98US-0099077.  
XX  
XX  
(LUDW-) LUDWIG INST CANCER RES.  
XX  
XX  
PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;  
XX  
XX  
DR WPT: 2000-256859/22.  
XX  
XX  
DR N-PSDB: AA296672.  
XX  
XX  
PT Isolated polypeptide used to treat subjects having a disorder  
characterized by expression of alternative open reading frame

PT macrophage-colony stimulating factor comprises 25 amino acid residue  
 PT sequence -  
 XX Claim 1; Page 64; 74pp; English.  
 XX  
 CC The present sequence represents a tumour rejection antigen precursor,  
 CC and is encoded by an alternative open reading frame (ORF) of human  
 CC macrophage colony stimulating gene. Peptides derived from the  
 CC alternative ORF of macrophage-colony stimulating factor, when presented  
 CC by an antigen presenting cell having a human leukocyte antigen (HLA)  
 CC class I molecule, effectively induce the activation and proliferation  
 CC of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic acids  
 CC derived from the alternate ORF of macrophage-colony stimulating factor  
 CC are useful for enriching selectively a population of T lymphocytes  
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder  
 CC characterized by expression of the polypeptide, and for identifying  
 CC functional variants and mimetics.  
 CC  
 XX  
 SQ Sequence 25 AA;  
 Query Match 100.0%; Score 130; DB 21; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-11;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MAGLPAAVGLSPGGEYHRGSGVYL 25  
 Db 1 maglpavvglspsgeyhrsgvyl 25  
 RESULT 2  
 AAY84265  
 ID AAY84265 standard; Peptide; 20 AA.  
 AC AAY84265;  
 XX  
 XX 12-JUL-2000 (first entry)  
 DT  
 DE Truncated macrophage colony stimulating factor tumour antigen.  
 XX  
 KW tumour rejection antigen; macrophage colony stimulating gene;  
 KW macrophage-colony stimulating factor; antigen presenting cell;  
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200013699-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PF 03-SEP-1999; 99WO-US20344.  
 XX  
 PR 04-SEP-1998; 98US-0099077.  
 XX  
 PA (LUDWIG -) INST CANCER RES.  
 XX  
 PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;  
 XX  
 DR WPI: 2000-256859/22.  
 DR N-PSDB; AA299675.  
 XX  
 XX Isolated polypeptide used to treat subjects having a disorder  
 PT characterized by expression of alternative open reading frame  
 PT macrophage-colony stimulating factor comprises 25 amino acid residue  
 PT sequence -  
 XX  
 PS Claim 3; Page 64; 74pp; English.  
 CC The present sequence represents a truncated tumour rejection antigen  
 CC precursor, and is encoded by a truncated alternative open reading frame  
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from  
 CC the alternative ORF of macrophage-colony stimulating factor, when  
 CC presented by an antigen presenting cell having a human leukocyte antigen

CC (HLA) class I molecule, effectively induce the activation and  
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic  
 CC acids derived from the alternate ORF of macrophage-colony stimulating  
 CC factor are useful for enriching selectively a population of T lymphocytes  
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder  
 CC characterized by expression of the polypeptide, and for identifying  
 CC functional variants and mimetics.  
 CC  
 XX  
 SQ Sequence 20 AA;  
 Query Match 81.5%; Score 106; DB 21; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-08;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MAGLPAAVGLSPGGEYHRG 20  
 Db 1 maglpavvglspsgeyhrg 20  
 RESULT 3  
 AAY84269  
 ID AAY84269 standard; Peptide; 15 AA.  
 AC AAY84269;  
 XX  
 XX 12-JUL-2000 (first entry)  
 DT  
 DE Peptide derived from macrophage colony stimulating gene alternative ORF.  
 XX  
 KW tumour rejection antigen; macrophage colony stimulating gene;  
 KW macrophage-colony stimulating factor; antigen presenting cell;  
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.  
 XX  
 OS Synthetic.  
 OS  
 PN WO200013699-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PF 03-SEP-1999; 99WO-US20344.  
 XX  
 PR 04-SEP-1998; 98US-0099077.  
 XX  
 PA (LUDWIG -) INST CANCER RES.  
 XX  
 PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;  
 XX  
 DR WPI: 2000-256859/22.  
 XX  
 XX Isolated polypeptide used to treat subjects having a disorder  
 PT characterized by expression of alternative open reading frame  
 PT macrophage-colony stimulating factor comprises 25 amino acid residue  
 PT sequence -  
 XX  
 PS Example 2; Page 40; 74pp; English.  
 CC The present sequence represents a peptide which is derived from a tumour  
 CC rejection antigen precursor encoded by an alternative open reading frame  
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from  
 CC the alternative ORF of macrophage-colony stimulating factor, when  
 CC presented by an antigen presenting cell having a human leukocyte antigen  
 CC (HLA) class I molecule, effectively induce the activation and  
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic  
 CC acids derived from the alternate ORF of macrophage-colony stimulating  
 CC factor are useful for enriching selectively a population of T lymphocytes  
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder  
 CC characterized by expression of the polypeptide, and for identifying  
 CC functional variants and mimetics.  
 CC  
 XX  
 SQ Sequence 15 AA;



Query Match 57.7%; Score 75; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 0.00043;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 AGLPAVGLSPGEQ 16  
 |||||  
 DB 1 eglpavglspgeqe 15

RESULT 4  
 AAY84266  
 ID AAY84266 standard; Peptide: 14 AA.  
 XX  
 AC AAY84266;  
 XX  
 DT 12-JUL-2000 (first entry)  
 XX  
 DE Peptide derived from macrophage colony stimulating gene alternative ORF.  
 XX  
 KW tumour rejection antigen; macrophage colony stimulating gene;  
 KW macrophage-colony stimulating factor; antigen presenting cell;  
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 PN MO200013699-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PF 03-SEP-1999; 99WO-US20344.  
 XX  
 PR 04-SEP-1998; 98US-0099077.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;  
 XX  
 DR WPI: 2000-256859/22.  
 XX  
 PT Isolated polypeptide used to treat subjects having a disorder  
 PT characterized by expression of alternative open reading frame  
 PT macrophage-colony stimulating factor comprises 25 amino acid residue  
 PT sequence -  
 XX  
 PS Claim 2; Page 39; 74pp; English.  
 XX  
 CC The present sequence represents a peptide which is derived from a tumour  
 CC rejection antigen precursor encoded by an alternative open reading frame  
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from  
 CC the alternative ORF of macrophage-colony stimulating factor, when  
 CC presented by an antigen presenting cell having a human leukocyte antigen  
 CC (HLA) class I molecule, effectively induce the activation and  
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic  
 CC acids derived from the alternate ORF of macrophage-colony stimulating  
 CC factor are useful for enriching selectively a population of T lymphocytes  
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder  
 CC characterized by expression of the polypeptide, and for identifying  
 CC functional variants and mimetics.  
 CC  
 XX  
 SQ Sequence 14 AA;

Query Match 55.4%; Score 72; DB 21; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.0011;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 LPNAVGLSPGEQ 17  
 |||||  
 DB 1 lpnavglspgeqey 14

RESULT 5  
 AAY84267  
 ID AAY84267 standard; Peptide: 13 AA.  
 XX  
 AC AAY84267;  
 XX  
 DT 12-JUL-2000 (first entry)  
 XX  
 DE Peptide derived from macrophage colony stimulating gene alternative ORF.  
 XX  
 KW tumour rejection antigen; macrophage colony stimulating gene;  
 KW macrophage-colony stimulating factor; antigen presenting cell;  
 KW human leukocyte antigen; CD8+ cytotoxic T lymphocyte.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 PN MO200013699-A1.  
 XX  
 PD 16-MAR-2000.  
 XX  
 PF 03-SEP-1999; 99WO-US20344.  
 XX  
 PR 04-SEP-1998; 98US-0099077.  
 XX  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 XX  
 PI Probst-Keppler M, Van Den Eynde B, Boon-Falleur T;  
 XX  
 DR WPI: 2000-256859/22.  
 XX  
 PT Isolated polypeptide used to treat subjects having a disorder  
 PT characterized by expression of alternative open reading frame  
 PT macrophage-colony stimulating factor comprises 25 amino acid residue  
 PT sequence -  
 XX  
 PS Example 2; Page 40; 74pp; English.  
 XX  
 CC The present sequence represents a peptide which is derived from a tumour  
 CC rejection antigen precursor encoded by an alternative open reading frame  
 CC (ORF) of human macrophage colony stimulating gene. Peptides derived from  
 CC the alternative ORF of macrophage-colony stimulating factor, when  
 CC presented by an antigen presenting cell having a human leukocyte antigen  
 CC (HLA) class I molecule, effectively induce the activation and  
 CC proliferation of CD8+ cytotoxic T lymphocytes. Polypeptide and nucleic  
 CC acids derived from the alternate ORF of macrophage-colony stimulating  
 CC factor are useful for enriching selectively a population of T lymphocytes  
 CC with CD8+ T lymphocytes. They are also useful for diagnosing a disorder  
 CC characterized by expression of the polypeptide, and for identifying  
 CC functional variants and mimetics.  
 CC  
 XX  
 SQ Sequence 13 AA;

Query Match 52.3%; Score 68; DB 21; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 0.0037;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 PAVVGLSPGEQ 17  
 |||||  
 DB 1 pavvglspegey 13

RESULT 6  
 AAY84268  
 ID AAY84268 standard; Peptide: 13 AA.  
 XX  
 AC AAY84268;  
 XX  
 DT 12-JUL-2000 (first entry)  
 XX  
 DE Peptide derived from macrophage colony stimulating gene alternative ORF.



PF 25-OCT-1999; 99MO-US24837.  
XX  
PR 02-NOV-1998; 98US-0184418.  
XX  
PA (UABR-) UAB RES FOUND.  
XX  
PI Hahn BH, Shaw GM, Gao F;  
XX  
DR WPI: 2000-365651/31.  
XX  
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency  
PT virus type 1 useful for detecting and treating AIDS comprises a  
PS specific nucleotide sequence -  
XX  
PS Claim 41; Fig 16; 131pp; English.  
XX  
CC The present invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection.  
XX  
SQ Sequence 191 AA:  
  
Query Match 39.6%; Score 51.5; DB 21; Length 191;  
Best Local Similarity 58.8%; Pred. No. 14;  
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
  
OY 9 GLSPGEQEHRC-GYGV 24  
DB 70 glpgeqwhlghvsl 86  
  
RESULT 9  
AAB69298  
ID AAB69298 standard; Protein: 192 AA.  
XX  
AC AAB69298;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE HIV-1 non-subtype B clone 962W751-3 vif protein.  
XX  
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env;  
KW vif; vpr; tat; rev; nef; vaccine.  
XX  
OS Human immunodeficiency virus type 1.  
XX  
PN MO200026416-A1.  
XX  
PD 11-MAY-2000.  
XX  
PE 25-OCT-1999; 99MO-US24837.  
XX  
PR 02-NOV-1998; 98US-0184418.  
XX  
PA (UABR-) UAB RES FOUND.  
XX  
PI Hahn BH, Shaw GM, Gao F;  
XX  
DR WPI: 2000-365651/31.  
XX  
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency  
PT virus type 1 useful for detecting and treating AIDS comprises a  
PS specific nucleotide sequence -  
XX  
PS Claim 41; Fig 16; 131pp; English.  
XX  
The present invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection.  
XX  
SQ Sequence 192 AA:  
  
Query Match 38.1%; Score 49.5; DB 21; Length 192;  
Best Local Similarity 58.8%; Pred. No. 27;  
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
  
OY 9 GLSPGEQEHRC-GYGV 24  
DB 71 glpgeqwhlghvsl 87  
  
RESULT 10  
AAB69290  
ID AAB69290 standard; Protein: 192 AA.  
XX  
AC AAB69290;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE HIV-1 non-subtype B clone 93BR020-1 vif protein.  
XX  
KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env;  
KW vif; vpr; tat; rev; nef; vaccine.  
XX  
OS Human immunodeficiency virus type 1.  
XX  
PN MO200026416-A1.  
XX  
PD 11-MAY-2000.  
XX  
PE 25-OCT-1999; 99MO-US24837.  
XX  
PR 02-NOV-1998; 98US-0184418.  
XX  
PA (UABR-) UAB RES FOUND.  
XX  
PI Hahn BH, Shaw GM, Gao F;  
XX  
DR WPI: 2000-365651/31.  
XX  
PT Novel genomic nucleic acids of non-subtype B human immunodeficiency  
PT virus type 1 useful for detecting and treating AIDS comprises a  
PS specific nucleotide sequence -  
XX  
PS Claim 41; Fig 16; 131pp; English.  
XX  
CC The present invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection.  
XX  
SQ Sequence 192 AA:  
  
Query Match 37.3%; Score 48.5; DB 21; Length 192;  
Best Local Similarity 58.8%; Pred. No. 38;  
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
  
OY 9 GLSPGEQEHRC-GYGV 24  
DB 11 glpgeqwhlghvsl 86

Db 71 gllpgerewhlgqvsl 87

# RESULT 11

AAE03295  
ID AAE03295 standard; Protein; 114 AA.

AC AAE03295;

DT 10-AUG-2001 (first entry)

XX Human gene 4 encoded secreted protein HNGMW50, SEQ ID NO:69.

XX Human: secreted protein; proliferative disorder; cancer; tumour;  
XX foetal abnormality; developmental abnormality; haematopoietic disorder;  
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
XX inflammation; allergy; neurological disorder; Alzheimer's disease;  
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
XX cardiovascular disorder; angiogenic disorder; kidney disorder;  
XX gastrointestinal disorder; pregnancy-related disorder;  
XX endocrine disorder; infection; wound healing; vulnerability;  
XX cell culture; chemotaxis; food additive; gene therapy;  
XX binding partner identification.

OS Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..16 /label= Signal\_peptide

XX Protein 17..114 /note= "Mature secreted protein"

XX W0200134800-A1.

XX 17-MAY-2001.

XX 08-NOV-2000; 2000WO-US30674.

XX 12-NOV-1999; 99US-0164750.

XX 30-JUN-2000; 2000US-0215128.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Edner R, Fiscella M, Wei P;

XX WPI; 2001-329085/34.

XX N-PSDB; AAD07708.

XX New nucleic acid molecules encoding human secreted proteins, used in

XX preventing, treating or ameliorating a disorder, e.g. Alzheimer's and

XX Parkinson's diseases and cancers -

XX Claim 11; Page 458; 530pp; English.

XX AAD07705-AAD07759 represent cDNAs corresponding to 19 human secreted  
XX protein genes, and AAE03292-AAE03346 represent the proteins they encode.  
XX AAE03347-AAE03375 represent human secreted protein fragments or variants.  
XX The genes and their secreted proteins are useful for preventing,  
XX treating or ameliorating conditions can be diagnosed by determining the  
XX amount of the new protein in a sample or by determining the presence of  
XX mutations in the new genes. Specific uses are described for each of the  
XX 19 genes, based on the tissues in which they are most highly expressed,  
XX and include developing products for the diagnosis or treatment of  
XX proliferative disorders, cancer, tumours, foetal and developmental  
XX abnormalities, haematopoietic disorders, diseases of the immune system,  
XX AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
XX allergies, neurological disorders (e.g., Alzheimer's disease,  
XX Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
XX skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
XX cardiovascular disorders, angiogenic disorders, kidney disorders,  
XX gastrointestinal disorders, pregnancy-related disorders, endocrine

CC disorders, and infections. The proteins can also be used to aid wound  
CC healing and epithelial cell proliferation, to prevent skin aging due to  
CC sunburn, to maintain organs before transplantation, for supporting cell  
CC culture of primary tissues, to regenerate tissues, to identify their  
CC cognate ligands or binding partners, and in chemotaxis, and can be used  
CC as a food additive or preservative to modify storage properties.  
CC Antibodies specific for a protein of the invention can be used in  
CC alleviating symptoms associated with the disorders mentioned above, and  
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
CC immunosorbent assay (ELISA). The present sequence represents a human  
CC secreted protein of the invention.

XX Sequence 114 AA;

XX Query Match 36.5%; Score 47.5; DB 22; Length 114;

XX Best Local Similarity 52.2%; Pred. No. 30;

XX Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 2 AGCPAYVGLSPGPO-EYHRCVG 23

Db 90 sglpaspqskgyryftrgr 112

# RESULT 12

AAW98485  
ID AAW98485 standard; Protein; 293 AA.

XX AAW98485;

XX 31-MAR-1999 (first entry)

XX H. pylori GHPD 997 protein.

XX GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;

XX peptic ulcer disease.

XX Helicobacter pylori.

XX W09843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

XX 01-APR-1997; 97US-0833457.

XX 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX14204.

XX Claim 8; Page 884-886; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHPD protein of the  
XX invention. The polypeptides can be used for preventing or treating  
XX Helicobacter infections, and gastroduodenal diseases associated with  
XX these infections, including acute, chronic, and atrophic gastritis, and  
XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
XX used for the production of antibodies. The products can also be used for  
XX detection and diagnosis.

SQ Sequence 293 AA;

Query Match 36.2%; Score 47; DB 19; Length 293;  
 Best Local Similarity 47.1%; Pred. No. 97;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 4 LPAVVGSLSPGEQYHRC 20  
 DB 206 lpaflgpnpyqafqerg 222

## RESULT 13

AAV58745  
 ID AAV58745 standard; Peptide: 46 AA.

AC AAV58745;

DT 25-APR-2000 (first entry)

DE Somatostatin antigenic peptide p2138b.

KW Pertussis toxin; helper T cell; Th epitope; feed additive;

KW growth promotion; somatostatin.

OS Chimeric - Bordetella pertussis.

OS Chimeric - Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..30

FT Peptide /note= "pertussis toxin Th epitope"

XX /note= "somatostatin"

PN W0996950-A1.

PD 29-DEC-1999.

PF 21-JUN-1999; 99WO-US13923.

PR 20-JUN-1998; 98US-0100415.

XX (UNBI-) UNITED BIOMEDICAL INC.

XX Wang CY;

DR MPI; 2000-160560/14.

PT New somatostatin helper T-cell epitope conjugate for raising  
 PT anti-somatostatin antibodies to enhance growth rate in animal by  
 PS reducing growth inhibitory activity of somatostatin

PS Claim 8; Page 41; 59pp; English.

CC The present sequence is that of somatostatin peptide immunogen  
 CC p2138b comprising a helper T cell (Th) epitope of the pertussis  
 CC toxin (see also AAV58741) joined via a dipeptide linker to  
 CC somatostatin. It is an example of immunogenic peptides of the  
 CC invention that are capable of inducing antibodies against  
 CC somatostatin, leading to the suppression of somatostatin levels,  
 CC thereby increasing growth rate and food conversion efficiency in  
 CC farm animals. The immunogenic peptides have a Th epitope linked  
 CC to a C- or N-terminal somatostatin or its peptide analogue, and may  
 CC also include an invasion domain as a general immune stimulator.

SO Sequence 46 AA;

Query Match 35.4%; Score 46; DB 21; Length 46;

Best Local Similarity 47.8%; Pred. No. 19;  
 Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 MAGLPAVVGSLSPGEQYHRC 23  
 DB 12 msflavrvhvskeegydyggag 34

## RESULT 14

ID AAB36208 standard; protein: 234 AA.

AC AAB36208;

DT 15-FEB-2001 (first entry)

DE Human immune system associated protein HISAP-6.

KW Human; immune system associated protein; HISAP-6; immune disorder;  
 KW infection; autoimmune disease; cancer.

OS Homo sapiens.

PN US6135941-A.

PD 24-OCT-2000.

PF 27-MAR-1998; 98US-0049672.

PR 27-MAR-1998; 98US-0049672.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Yue H, Lal P, Corley NC, Guegler KJ, Baughn MR;

PI Hillman JL, Au-Young J;

DR MPI: 2001-030926/04.

DR N-PSDB; AAC6524.

PT New human immune system associated proteins (HISAP) and polynucleotides  
 PT encoding the HISAP, useful for diagnosing, treating or preventing  
 PS immune or cell proliferative disorders or infections

PS Claim 1; Column 59-60; 54pp; English.

CC The present invention provides the coding and protein sequences for a  
 CC number of human immune system associated proteins (HISAPs). These can be  
 CC used in the diagnosis and treatment of various autoimmune disorders,  
 CC infections and cell proliferation diseases. The diseases include AIDS,  
 CC adult respiratory distress syndrome, anaemia, asthma, atherosclerosis,  
 CC Crohn's disease, irritable bowel syndrome, multiple sclerosis, myasthenia  
 CC gravis, osteoarthritis, rheumatoid arthritis, scleroderma, systemic lupus  
 CC erythematosus, arteriosclerosis, cirrhosis and cancer.

SO Sequence 234 AA;

Query Match 35.4%; Score 46; DB 22; Length 234;  
 Best Local Similarity 60.0%; Pred. No. 1,1e+02;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAGLPAVVGSLSPGRQ 15  
 DB 24 mtgspavvslspgr 38

## RESULT 15

AAV16108

ID AAV16108 standard; protein: 300 AA.

AC AAV16108;

DT 05-AUG-1999 (first entry)

DE A formate transport associated protein, PurU.

KW Formate transport associated protein; FTAPl; formate production;  
 KW gram-positive microorganism; formate transport; formate utilization;  
 KW formate cycling; FTAPl; PurU; FMD; large scale fermentation.

```

XX OS Bacillus subtilis.
XX PN WO9927107-A2.
XX PD 03-JUN-1999.
XX PF 19-NOV-1998; 98WQ-US24873.
XX PR 20-NOV-1997; 97GB-0024627.
XX PA (GEMV ) GENENCOR INT INC.
XX PI Diaz-Torres M, Miller B;
XX WPI: 1999-357843/30.
XX DR N-PSDB: AAX60111.
XX PS Regulation of formate production in gram-positive microorganisms
XX PT
XX PS Disclosure: Fig 3A-C; 58pp; English.
XX CC The present sequence represents a formate transport associated protein,
XX CC puru. The specification describes a method for modulating formate
XX CC production levels in gram-positive microorganisms. Formate production
XX CC levels in gram-positive microorganisms can be regulated by the modulation
XX CC of four Bacillus proteins which appear to be associated with formate
XX CC transport, utilization and cycling. These proteins are FTAP1, FTAP2, puru
XX CC and FMD. This also allows for increased yields of recombinant proteins
XX CC being produced by the microorganism. The methods of the invention can
XX CC be used to increase recombinant protein production in gram-positive host
XX CC microorganisms. The protein produced by a recombinant gram-positive host
XX CC cell comprising a FTAP1, FTAP2, puru and FMD will be secreted into the
XX CC cell culture, from which it can be purified. The host cell can also be
XX CC used to produce FTAP1, FTAP2, puru and FMD proteins, preferably under
XX CC large scale fermentation conditions. FTAP1, FTAP2, puru and FMD
XX CC polynucleotides can be used to detect the presence of gram-positive
XX CC microorganism polynucleotide homologues through hybridization techniques
XX CC and PCR. The polynucleotides also serve as a source of primers and
XX CC probes.
XX SQ Sequence 300 AA;

```

```

Query Match 35.4%; Score 46; DB 20; Length 300;
Best Local Similarity 47.1%; Pred. No. 1.4e+02;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY 4 LPAVVGSLSPGEQEXYHRG 20
| | | : | : | : | |
Db 212 lpatiganpykkrayerg 228

```

Search completed: September 3, 2002, 15:22:10  
 job time: 378 sec